HET-NETs 2010
ISBN XXX–XXX
pp. xx–xx

Moderate deviations of retransmission buffers over a wireless fading channel

KOEN DE TURCK [a] MARC MOENECLAEY[a] SABINE WITTEVRONGEL[a]

[a]Department of Telecommunications and Information Processing
Ghent University, Belgium
kdeturck@telin.ugent.be

Abstract: We study the buffer performance of retransmission protocols over a wireless link. The channel gain is modeled as a discrete-time complex Gaussian model. The advantage of this channel model over simpler alternatives (such as finite-state Markov models) is that the correspondence with the physical situation is more transparent. In order to keep the performance analysis of the buffer content tractable, we opt for heavy-traffic and moderate deviations scalings. We add some notes on the optimal selection of the signal to noise ratio.
Keywords: Moderate deviations, wireless channels, fading models.

1. Introduction

The popularity of wireless telecommunications is increasing rapidly. However, apart from the obvious advantages over wired networks, such as increased user mobility and easier deployment, wireless communications also have a number of drawbacks. For example, the dependence on batteries requires a more careful energy management. Secondly, the presence of fading and interference is also particular to wireless links and may cause a severe degradation of performance.

In this paper, we look at the performance loss due to fading. This loss manifests itself in a reduced throughput of the link, but other performance metrics may be severely affected as well, such as the mean packet delay, the overflow probability of the buffer at the transmitter's side or the delay jitter. This is by no means a new topic; we refer to [8, 7, 3] as a sample of how this problem has been tackled, some of the papers focus on throughput only, others on the complete buffer performance. The novelty of this paper resides in the fact that we combine a couple of elements in a way that has not been done before, which leads to an elegant analysis. Firstly, we make use of a complex Gaussian process as a model of the channel gain. This type of model is more popular in communications theory circles, than for queueing analyses, because a direct computation of the buffer content distribution is too resource-intensive to be useful. For this reason, researchers interested in the buffer content have hitherto focused on finite-state Markov models. The undeniable advantage of Gaussian models however is that the correspondence with physical parameters is transparent: metrics such as the SNR (signal-to-noise-ratio) and coherence time feature directly.

During recent years, scaling has become a respected and full-fledged analysis tool. When confronted with a problem for which a direct computation is very costly or plainly impossible, the probabilist might opt to scale the original problem in such manner that a much simpler model arises, one in which the salient features of the original model are retained, but other stochastic fluctuations get filtered away. We look at two scaling methods in particular, namely heavy-traffic and moderate deviations. Heavy-traffic analysis is easily the oldest scaling method known in queueing theory. Kingman was the first to exploit the deep link between queueing systems operating under the limit load $\rho \to 1$ and diffusion processes. Moderate deviations do not have such a long history. Its promise is to combine the strong points of large deviations and heavy traffic methods. Essentially, it is a rigourous way of looking at tails of asymptotically Gaussian processes. We are indebted to the scaling approach taken in [11, 10].

The structure of the paper is as follows. In section 2., we detail the channel and the buffer model; in section 3., we review the moderate deviations and heavy traffic scalings and apply them to the model at hand. We look at some numerical examples in section 5. and finally, we draw conclusions in section 6.

2. Model

Consider a wireless station (the transmitter) delivering data packets to another wireless station (the receiver). Time is considered to be slotted, where the duration of a slot corresponds to the transmission time of a data packet of length L bits. The transmission buffer has room for B data packets. The channel over which the information is sent is subject to fading, which we model as follows. The channel gain h_t during slot $t \in \mathbb{N}$ forms a discrete-time complex Gaussian process. We assume wide-sense stationarity, and moreover $\mathrm{E}\,\mathsf{h}_t = 0$. The process is thus characterized completely by an autocorrelation function r_t:

$$r_t \doteq \mathrm{E}(\mathsf{h}_s^* \mathsf{h}_{t+s}) = \mathrm{E}(\mathsf{h}_0^* \mathsf{h}_t),$$

where z^* denotes the complex conjugate of z.

The process $\{\mathsf{h}_t\}$ can also be characterized as filtered white noise. Indeed, consider a sequence u_t of independent and identically distributed (iid) complex normal variables with zero mean and unit variance, and a filter bank with parameters g_t such that:

$$\mathsf{h}_t = \sum_s g_s \mathsf{u}_{t-s}.$$

The two representations are in fact equivalent. A popular choice in this case is the Butterworth filter. In this paper, however, we do not further elaborate on the filter representation of the channel process.

Two choices of r_t are particularly popular. The so-called Jakes' model is perhaps the most well-known choice. It was derived from theoretical considerations, and expresses r_t in terms of a Bessel function of the first kind:

$$r_t = J_0(2\pi f_d t),$$

where f_d is the Doppler frequency. There is a simple relation between the the Doppler frequency f_d, the carrier frequency f_c and the velocity of the receiver:

$$f_d = \frac{v}{c} f_c,$$

where c denotes the speed of light. This shows the strong influence of the carrier frequency and the velocity on the nature of the fading process. The Doppler frequency also corresponds to the cut-off frequency of the filter.

The other popular form for the autocorrelation function is a Gaussian form:

$$r_t = \exp\left(-\frac{t^2}{2\alpha^2}\right), \tag{1}$$

where α is a form factor regulating the 'width' of the autocorrelation function. We can relate this to the Doppler frequency by determining the cut-off frequency in the frequency domain. A Gaussian function with form factor α in the time domain is mapped unto a Gaussian function with form factor α^{-1} in the frequency domain. Some manipulations yield the following formula for the n-dB cut-off frequency:

$$f_d = \sqrt{\frac{n}{5}\log 10}\,\alpha^{-1}. \tag{2}$$

Our overview of the channel model is completed by the link between the channel gain and the transmission error process. The bit error probability is a function of the channel gain h as follows:

$$p_b(h) = Q\left(\sqrt{\frac{2E_b}{N_0}|h|^2}\right),$$

where $\frac{E_b}{N_0}$ denotes the SNR and $Q(x)$ denotes the error function; it is equal to the probability that a normal random variable with zero mean and unit variance is larger than x. The packet error probability $p(h)$ is the probability that at least one bit of the packet is incorrect:

$$p(h) = 1 - (1 - p_b(h))^L.$$

Let $\{\mathsf{c}_t\}$ denote the transmission process: c_t is equal to 1 when the transmission during slot t is successful and 0 otherwise. We have that

$$\mathsf{c}_t = \mathsf{i}_{1-p(\mathsf{h}_t)},$$

where i_q denotes a Bernoulli random variable with success probability q.

Let $\{\mathsf{a}_t\}$ be the random process of the number of packet arrivals during slot t. A natural class of arrival processes for this kind of analysis is that they are asymptotically Gaussian under the scaling we are considering. We will provide more details as we go along. Stationarity is another natural condition that we impose throughout this paper. Let $\lambda \doteq \mathrm{E}\,\mathsf{a}_0; \mu \doteq \mathrm{E}\,\mathsf{c}_0$. The load of the system is defined as $\rho = \lambda/\mu$.

In this paper, we look at the transmitter buffer performance, with the so-called 'ideal ARQ' (ARQ stands for automatic repeat request) protocol: packets are retransmitted until they are received correctly, (until $\mathsf{c}_t = 1$) with the assumption that there is no feedback delay. That is, the transmission status of a packet is directly known. The scalings that we consider in this paper involve letting the load approach 1, and under such conditions ARQ with non-zero feedback delay converges to ideal ARQ. The queue content process $\{\mathsf{q}_t\}$ is formulated in terms of the arrival and transmission processes, by means of the well-known Lindley recursion:

$$\mathsf{q}_{t+1} = [\mathsf{q}_t + \mathsf{a}_t - \mathsf{c}_t]_0^B.$$

where $[x]_0^B \doteq \max(0, \min(B, x))$.

3. Scalings

We obtain asymptotic results on the queue content distribution by appropriately scaling the arrival and transmission streams. In this context, it is customary to define the net-input process $\mathsf{w}_t \doteq \mathsf{a}_t - \mathsf{c}_t$. Even within the class of scaling methods (which are by themselves already approximative) we have to be careful as to which methods offer good approximations for a reasonably low computational effort.

3.1. Fast-time scaling

We consider a set of scalings that involve speeding up the net-input process. Let $\mathsf{w}^{\otimes L}$ denote the net-input process sped up by a factor L:

$$\mathsf{w}_t^{\otimes L} \doteq \sum_{s=tL}^{(t+1)L-1} \mathsf{w}_s.$$

Let us look at a family of scalings of the form:

$$\hat{\mathsf{w}} = L^{(1-\beta)/2}(L^{-1}\mathsf{w}^{\otimes L} - (\lambda - \mu)\mathsf{1}). \tag{3}$$

where $\beta \in [0, 1]$ and $\mathsf{1}$ denotes a constant process and equal to 1. For $\beta = 0$, we get the so-called central limit scaling, whereas $\beta = 1$ is known as the large deviations scaling, (which is essentially the same as the scaling used for the law of large numbers).

Let us first have a look at the central limit scaling $\beta = 0$. Under mild conditions (typically the existence of the first two moments, and some mixing condition), the scaled process converges to a (discrete sample of) Brownian motion with zero drift and diffusion parameter V^{w}:

$$V^{\mathsf{w}} = \lim_{t\to\infty} \mathrm{Var}[\sum_{s=1}^{t} \mathsf{w}_s].$$

The queue content process under this scaling converges to a reflected Brownian motion with drift $\lambda - \mu$, diffusion parameter σ_{w}^2, and boundaries at 0 and B. This leads to a couple of simple performance formulae:

$$\mathrm{E}[\mathsf{q}] \approx \frac{V^{\mathsf{w}}}{2(\mu-\lambda)}, \tag{4}$$

and

$$\Pr[\mathsf{q} \geq b] \approx \exp\left(-\frac{2b(\mu-\lambda)}{V^{\mathsf{w}}}\right) \tag{5}$$

Large and moderate deviations scalings are less useful for this application. They involve computing a rate function, which for the transmission process at hand is either computationally complex (in the large deviations case), or leads only to the asymptotic formula (5) of the central limit result (in the moderate deviations case). Indeed, for the large deviations case, computations center around the scaled cumulant generating function (scgf) $\Lambda(\theta)$:

$$\Lambda(\theta) = \lim_{T\to\infty} \frac{1}{T} \log \mathrm{E}[\exp(\theta \mathsf{w}_0^{\otimes T})],$$

from which the rate function can be obtained by a Legendre-Fenchel transform. The computation of $\Lambda(\theta)$ requires the evaluation of a high-dimensional integral (with the dimension tending to infinity). This can only be solved by costly Monte-Carlo techniques, and as we set out to find easy to compute performance formulae, we will not pursue this path further. We note that the moderate deviations limit in fact corresponds to a second-order

3.2. Many flows scalings

We now look at a scaling that preserves the time-covariance structure of the original net-input process: instead of speeding up this process, we denote by $\mathsf{w}^{\oplus L}$ an aggregate of L independent copies of the same net-input process. The family of scalings has now the following form:

$$\hat{\mathsf{w}} = L^{(1-\beta)/2}(L^{-1}\mathsf{w}^{\oplus L} - (\lambda-\mu)1), \tag{6}$$

again for $\beta \in [0,1]$. In the central limit scale $\beta = 0$, the scaled process now converges to a Gaussian process (not necessarily Brownian motion) with the same drift and covariance structure as the original net-input process w. Although the queue-content process also converges to a Gaussian process, it is generally difficult to derive closed-form performance metrics for it. This is why we resort to moderate deviations in this case. Under some mild conditions, the scaled process satisfies a moderate deviations principle (MDP) for $\beta \in (0,1)$ with rate function I_t:

$$\lim_{L\to\infty} L^{-\beta} \log \Pr[\hat{\mathsf{w}} \in \hat{S}] \asymp -\inf_{t>0} \inf_{\hat{\mathsf{x}}\in\hat{S}} I_t(\hat{x}), \tag{7}$$

where $I_t(x)$ is equal to

$$I_t(x) = \sup_{\theta \in \mathbb{R}^t} \theta^\mathsf{T} x - \frac{1}{2}\theta^\mathsf{T} C_t \theta, \tag{8}$$

with C_t the covariance matrix of the net-input process (with dimension $t \times t$):

$$[C_t]_{ij} = \gamma_{|i-j|} = \mathrm{Cov}[\mathsf{w}_i, \mathsf{w}_j]. \tag{9}$$

The tail asymptotics of the queue content process are given by [11]:

$$\log \Pr[\mathsf{q} \geq b] \asymp -I, \tag{10}$$

where

$$I = \inf_{t \geq 0} \frac{(b + (\mu - \lambda)t)^2}{2V_t}. \tag{11}$$

We detail in the next subsection how to compute the variance function V_t, which is defined as follows: $V_t = \sum_{i,j} [C_t]_{ij}$. One can also use the 'refined asymptotics' of the Bahadur-Rao type [12]:

$$\Pr[\mathsf{q} \geq b] \approx \frac{1}{\theta^* \sqrt{2\pi V_{t^*}}} e^{-I}, \tag{12}$$

where t^* is the t that minimizes (11), and $\theta^* = (b + (\mu - \lambda)t)/2V_{t^*}$.

3.3. Computing the covariance structure

In this section, we show how to compute the function V_t that appears in the asymptotic performance measures of the previous section. First, note that the net-input process is the sum of two independent processes: the arrival process and the transmission process, which means that V_t can be split up likewise:

$$V_t = V_t^\mathsf{a} + V_t^\mathsf{c}. \tag{13}$$

For the arrival process, we opt in this paper for the parsimonious fractional Brownian process, which has three parameters: a drift λ, a diffusion parameter σ^2 and a Hurst parameter H, where $H \in (0,1)$. For $H = \frac{1}{2}$, we have the standard Brownian motion with independent increments, In case of $H < \frac{1}{2}$ the increments are negatively correlated, and positively correlated for $H > \frac{1}{2}$. We have that $V_t^\mathsf{a} = \sigma^2 t^{2H}$.

The function V_t^c of the transmission process can be found via the auxiliary sequence γ_t:

$$\begin{aligned} \gamma_t &\doteq \mathrm{E}[(\mathsf{c}_0 - \mu)(\mathsf{c}_t - \mu)] \\ &= \mathrm{E}[(1 - p(\mathsf{h}_0) - \mu)(1 - p(\mathsf{h}_t) - \mu)]. \end{aligned} \tag{14}$$

where the last transition is due to the definition of $\mathsf{i}(.)$. The computation of this sequence is best done numerically. The computational complexity is relatively minor, however: for each

t we must evaluate a four-dimensional integral (recall that h_t are complex-valued random variables, thus yielding two dimensions each). Sequences V_t^{c} and γ_t are related as follows:

$$V_t^{\mathsf{c}} = \sum_{i=0}^{t-1} \sum_{j=0}^{t-1} \gamma_{|i-j|}. \tag{15}$$

The asymptotic variance, which plays a central role in fast-time scalings, is equal to the limit $V = \lim_{t\to\infty} V_t/t$. Note that this limit may not exist, for example for fractional Brownian with Hurst parameter $H > \frac{1}{2}$.

4. Optimal control

The transmission over wireless channels poses a challenging control problem to the designer of wireless networks: which level of the SNR represents the optimal trade-off between quality-of-service and energy consumption ? The parsimonious performance formulae presented in the previous section offer a feasible path to the static optimization. Indeed, assume given as a QoS constraint that the overflow probability must be smaller than P. The SNR influences the transmission rate μ and the variability of the transmission process V_t^{c}. The buffer size b influences the overflow probability. The control problem is thus as follows:

Find the minimal buffer size b and SNR such that:

$$-\log P > \inf_{t\geq 0} \frac{(b + (\mu(\text{SNR}) - \lambda)t)^2}{2(V_t^{\mathsf{a}} + V_t^{\mathsf{c}}(\text{SNR}))}.$$

One can also adapt the SNR dynamically according to the perceived current channel and traffic state. This dynamic optimization problem is a lot harder, and is the subject of future research.

5. Some numerical results

Consider a scenario in which a transmitter sends data packets of $L = 10000$ bits to a receiver over a wireless channel subject to fading. The duration of a packet transmission is 5 ms. The carrier frequency f_c of the transmission is $1Gz$, and the receiver moves relative to the sender with a velocity of v. In the first pair of figures, we show autocorrelation function of the channel gain for different velocities and for the Gaussian and Jakes' model respectively. Note that the manner in which the two models decay is completely different. We also show the corresponding autocorrelation function γ_t of the transmission process c_t for the same scenarios in figure 2. Note that again, for a Jakes' model there are a lot of small bumps after the main bump, whereas for the Gaussian model there is only one bump. Although the bumps appear small they have an considerable influence on the function V_t.

Next, we turn our attention to the buffer performance proper. We plot the tail probabilities of the buffer occupancy for different velocities in the left subplot of figure 3. We see that the velocity has a huge effect on the buffer performance. In the right subplot, we look at the log-probability of the buffer exceeding a certain level ($b = 80$) versus the velocity v. We see that above some speed the influence gets minimal.

In the last figure, we show that when the arrival source is really bursty (Hurst parameter $H = 0.7$, signifiying a large positive correlation), the performance of the buffer deteriorates to the extent that the influence of the fading channel is hardly seen.

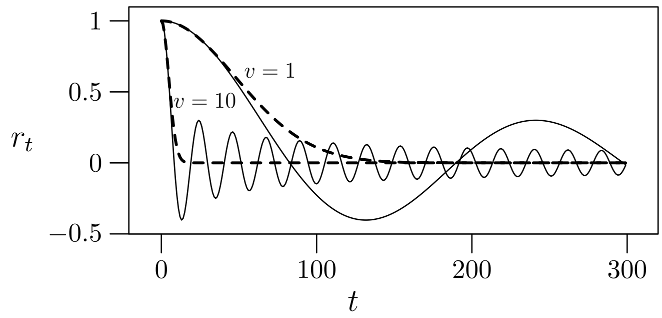

Fig. 1: Channel autocorrelation functions r_t with ‘Gaussian’ form (dashed lines) and Besselian form (Jakes’ model; full lines) for different values of v(in km/h).

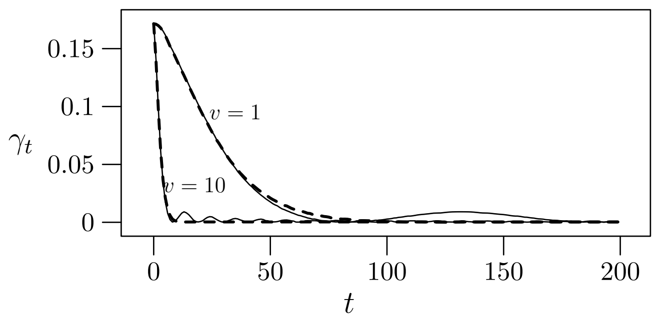

Fig. 2: Transmission autocorrelation function γ_t for different values of v. Gaussian-form models are in dashed lines; Jakes’ models are in full lines. Values of the other parameters are: $f_c = 1$ GHz; $T_p = 5$ms;$L = 10000$; SNR=14dB.

6. Conclusion

We studied the moderate deviations asymptotics of a retransmission buffer over a wireless fading channel. We found easy to evaluate performance formulae that link important physical parameters such as signal-to-noise ratios, coherence time and so on. The most important conclusions are: (1) the throughput alone does not suffice to characterize the buffer performance, (2) the lower the velocity of the receiver the worse the buffer performs (3) the effects of the fading channel might be swamped by really bursty (or ‘Hursty’) arrival sources, especially when the packet error probability is reasonably low, and (4) Gaussian and Jakes’ fading models give different tail behavior.

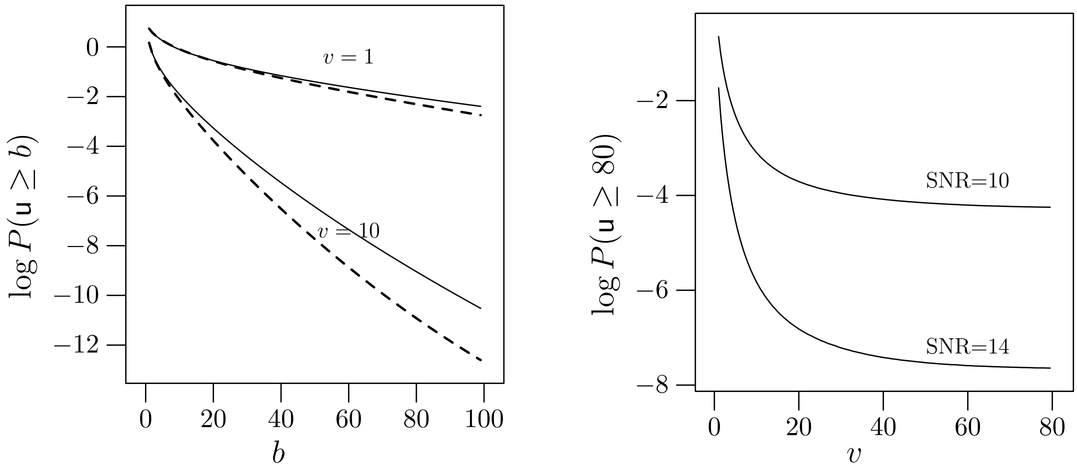

Fig. 3: Left: Log probabilities of the buffer content u for different values of v; Gaussian-form models are in dashed lines; Jakes' models are in full lines. Right: $\log P(\mathsf{u} \geq 80)$ versus the velocitity v of the receiver, for a Gaussian-form model. Values of other parameters are: SNR=14dB;$H = 0.5$;$V^{\mathsf{a}} = 0.1$.

References

[1] Bhunia, C.T.: ARQ - Review and modifications. IETE Technical Review **18** (2001) 381–401

[2] Towsley, D., Wolf, J.K.: On the statistical analysis of queue lengths and waiting times for statistical multiplexers with ARQ retransmission schemes. IEEE Transactions on Communications **25** (1979) 693–703

[3] Konheim, A.G.: A queueing analysis of two ARQ protocols. IEEE Transactions on Communications **28** (1980) 1004–1014

[4] De Munnynck, M., Wittevrongel, S., Lootens, A., Bruneel, H.: Queueing analysis of some continuous ARQ strategies with repeated transmissions. Electronics Letters **38** (2002) 1295 – 1297

[5] Towsley, D.: A statistical analysis of ARQ protocols operating in a non-independent error environment. IEEE Transactions on Communications **27** (1981) 971–981

[6] De Vuyst, S., Wittevrongel, S., Bruneel, H.: Delay analysis of the Stop-and-Wait ARQ scheme under correlated errors. Proceedings of HET-NETs 2004, Performance Modelling and Evaluation of Heterogenous Networks (26-28 July 2004, Ilkley, West Yorkshire, UK), p. 21/1–21/11

[7] R. Fantacci. Queueing analysis of the Selective Repeat automatic repeat request protocol wireless packet networks. *IEEE Trans. Veh. Technol*, 45:258–264, 1996.

[8] Kim, J.G., Krunz, M.: Delay analysis of selective repeat ARQ for transporting Markovian sources over a wireless channel. IEEE Transactions on Vehicular Technology **49** (2000) 1968–1981

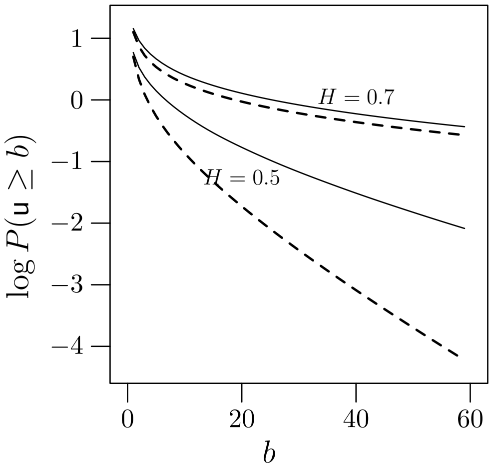

Fig. 4: A plot of $\log P(\mathsf{u} \geq b)$ versus b for an SNR of 10dB (dashed lines) and of 14dB (full lines) for two values of the Hurst parameter. Values of the other parameters are: $V_\mathsf{a} = 1$; $v = 5$km/h.

[9] Gilbert, E.N.: Capacity of a burst noise channel. The Bell System Technical Journal **39** (1960) 1253–1265

[10] Ayalvadi Ganesh, Neil O'Connell, Damon Wischik: Big Queues. Springer Verlag Berlin (2004)

[11] Damon Wischik: Moderate deviations in queueing theory (submitted for publication)

[12] Montgomery, M., DeVeciana, G.: On the relevance of time scales in performance oriented traffic characterizations. Proceedings of INFOCOM 1996.